Case-study 1

Analysing global video game sales

Analysing global video game sales

Project Overview

GameCo is a new videogames company that wants to use data to inform the development of new games. My task was to perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's games might fare in the market.

Key Questions and Objectives

- Are certain types of games more popular than others?
- What other publishers will likely be the main competitors in certain markets?
- Have any games decreased or increased in popularity over time?
- How have their sales figures varied between geographic regions over time?

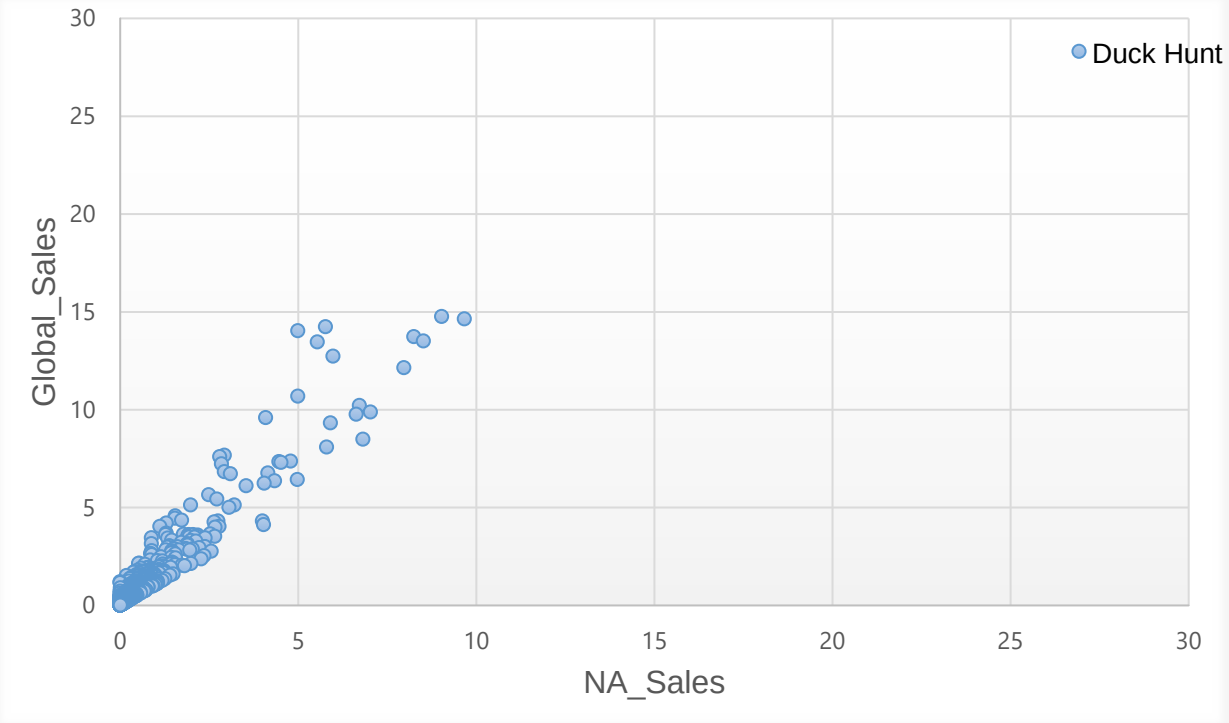
Technical information

- Software used: Microsoft Excel
- The data set covers historical sales of video games across different platforms, genres, and publishing studios. It contains 16.598 rows, and it was sourced from the website vgchartz.com.

Analysing global video game sales

Tasks performed

- Used descriptive statistics and visualizations to perform exploratory data analysis.
- Checked and fixed duplicates and missing values
- Used pivot tables to summarize the data, making it more usable and insightful.
- Filtered the data to look at specific subsegments.
- Created new variables that expand the meaning of data.
- Worked through a series of hypothesis derived from the data.
- Created visualization, such as bar and column charts, box and whisker plots, and scatter plots, to answer business questions.



Analysing global video game sales

Result:

1. Created a final presentation to communicate the findings, conclusions and recommendations to stakeholders.

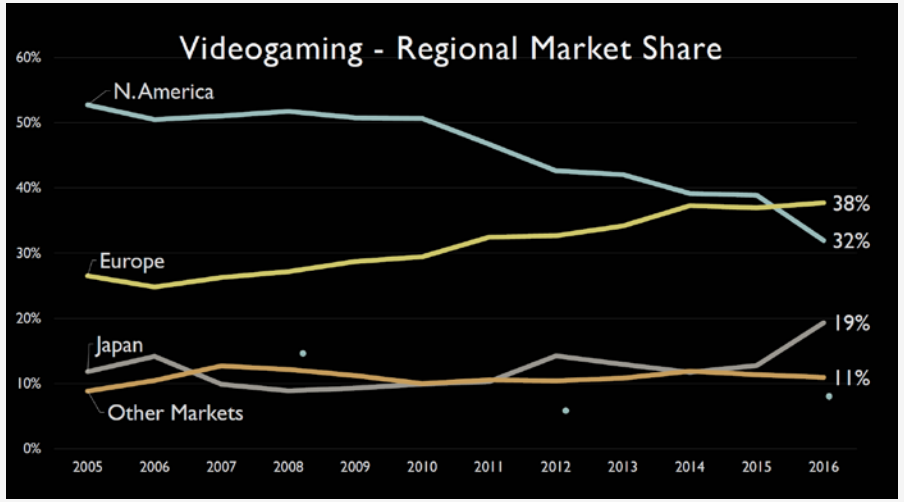
Main conclusions:

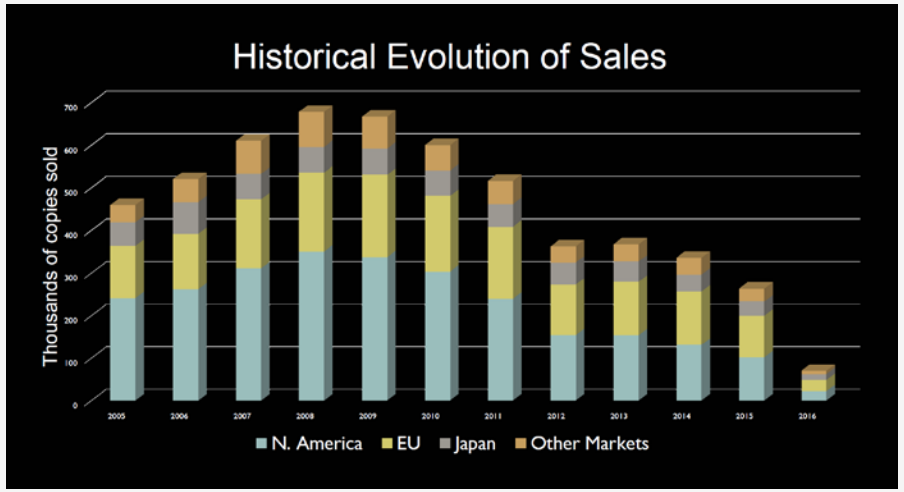
1. During the last 12 years, on average, NA sales represented 46% of global sales, while the EU and the JP market accounted for 31% and 12% respectively.

2. Between 1980 and 2016, more than 85% of games were produced for the 12 most popular platforms.

3. From the 579 publishers included in our database, 15 of them account for more than 80% of global sales. Nintendo is the leader with 20% of the market!

4. 2008 was the year most games were sold worldwide. It has decrease since then. The amount of games sold globally in 2015 was 61% lower than in 2008!





Case-study 2

Preparing for flu season in the U.S.

Preparing for flu season in the U.S.

Project Overview

The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients.

Objective

To help a medical staffing agency to plan for the influenza season, a time when additional staff are in high demand. My job was to examine trends in influenza and understand how they can be used determine when to send staff, and how many, to each state.

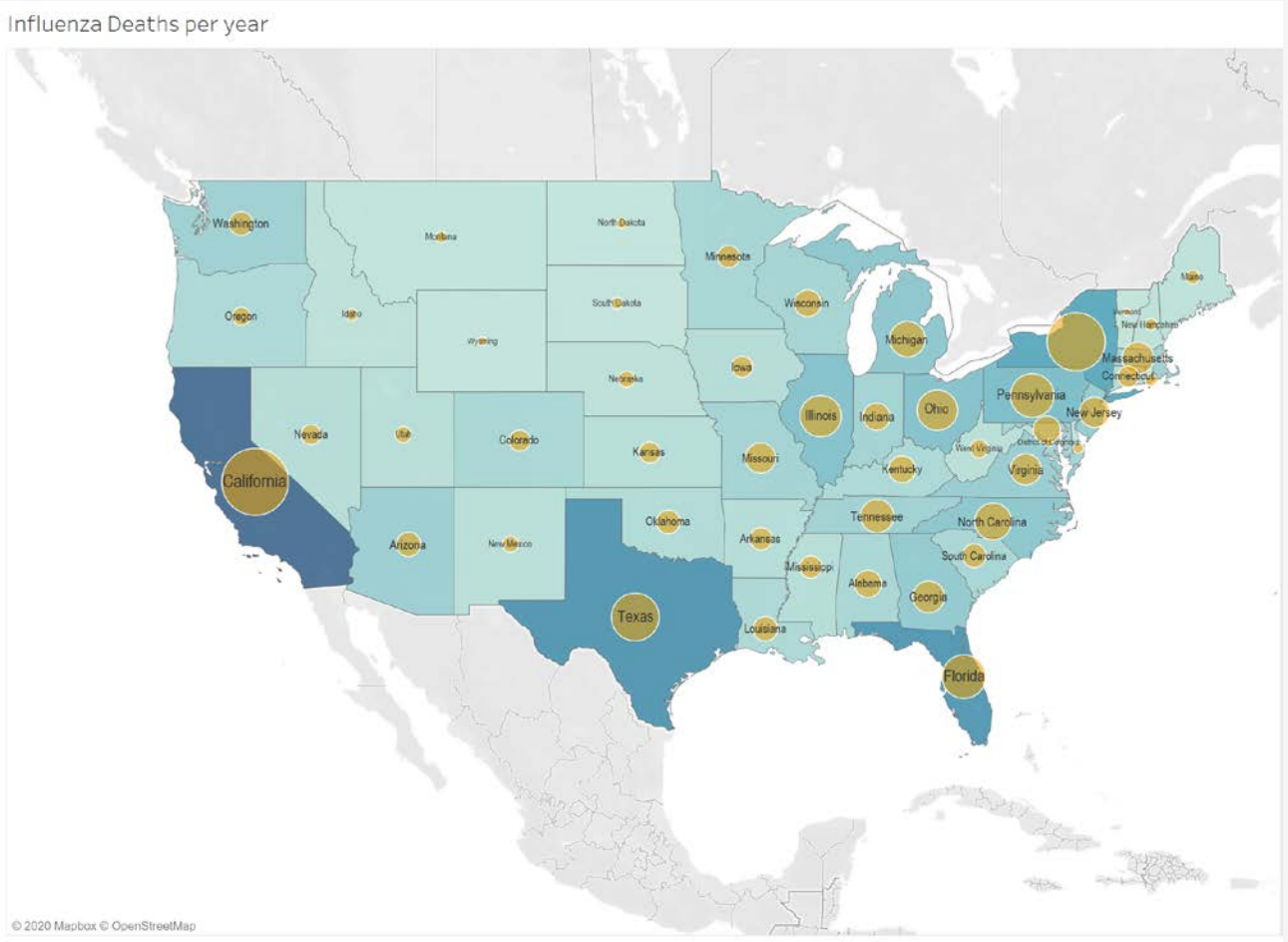
Technical information

- Software used: Excel and Tableau
- The following data sets were used:
  - Influenza deaths by geography, time, age, and gender (source: CDC)
  - Population data by geography (US Census Bureau)
  - Counts of influenza laboratory test results by state (CDC)
  - Survey of flu shot rates in children (CDC)

Preparing for flu season in the U.S.

Tasks performed

- Formulated the research hypothesis.
- Sourced the right data.
- Checked and solved data integrity issues in the data sets.
- Integrated data from two sources into one cohesive data set using data transformations.
- Conducted regression analysis to test hypothesis.
- Elaborated an interim report to share the findings of my analysis.
- Created interactive visualizations in Tableau:
  - ✓ Column charts
  - ✓ Population pyramids
  - ✓ Histograms
  - ✓ Time-series forecast
  - ✓ Box & whiskers,
  - ✓ Bubble charts
  - ✓ Spatial Analysis



Preparing for flu season in the U.S.

Results:

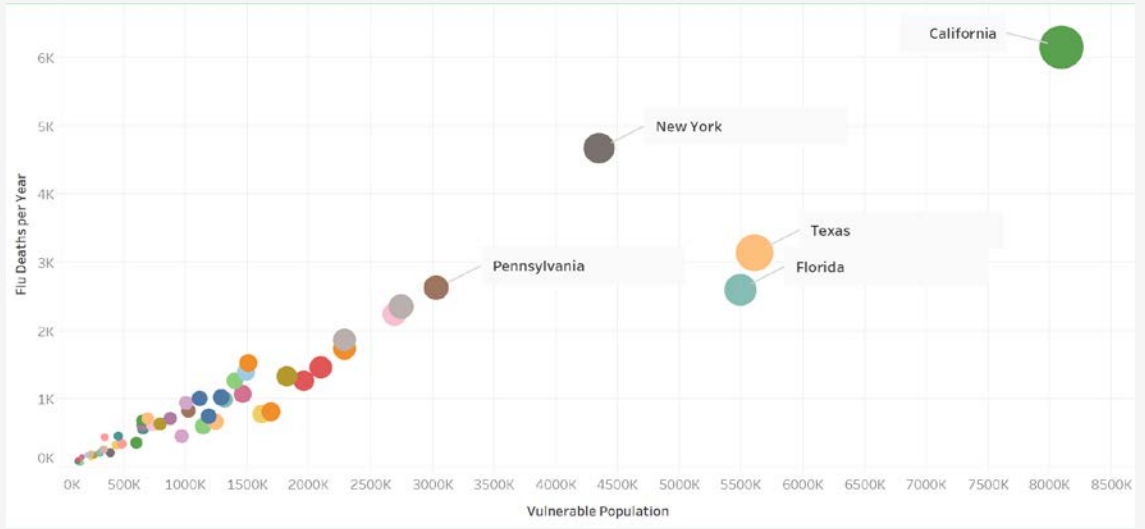
1. Created a storyboard to communicate the findings using Tableau Business Intelligence and Analytics Software

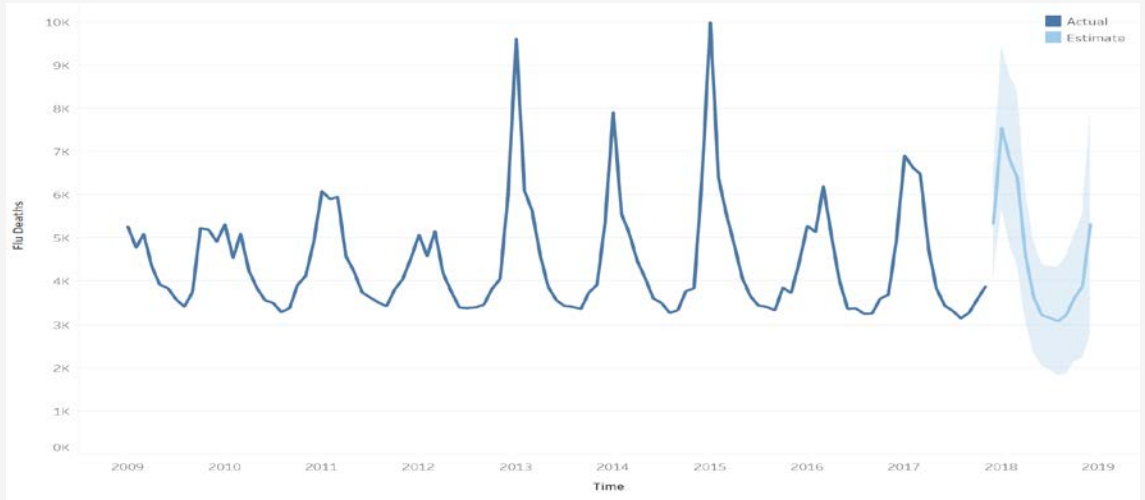
2. Created a video presentation to guide the audience through the storyboard.

Main Conclusions:

1. The States with more population, and therefore more vulnerable population, have more deaths due to the flu. Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital, determining the need of medical personnel.

2. The flu season in the U.S.A is from November to May. The precise timing and severity varies between States.





Case-study 3

Answering business questions for an online video rental company

Answering business questions for an online video rental company

Project Overview

Rockbuster Stealth LLC is a movie rental company that used to have stores around the world and is now planning to launch an online video rental service. My job was to load Rockbuster's data into a RDBMS, perform data analysis to answer business questions and present the results.

Key Questions and Objectives

- Which movies contributed the most/least to revenue gain?
- What was the average rental duration for all videos?
- Which countries are Rockbuster customers based in?
- Where are customers with a high lifetime value based?
- Do sales figures vary between geographic regions?

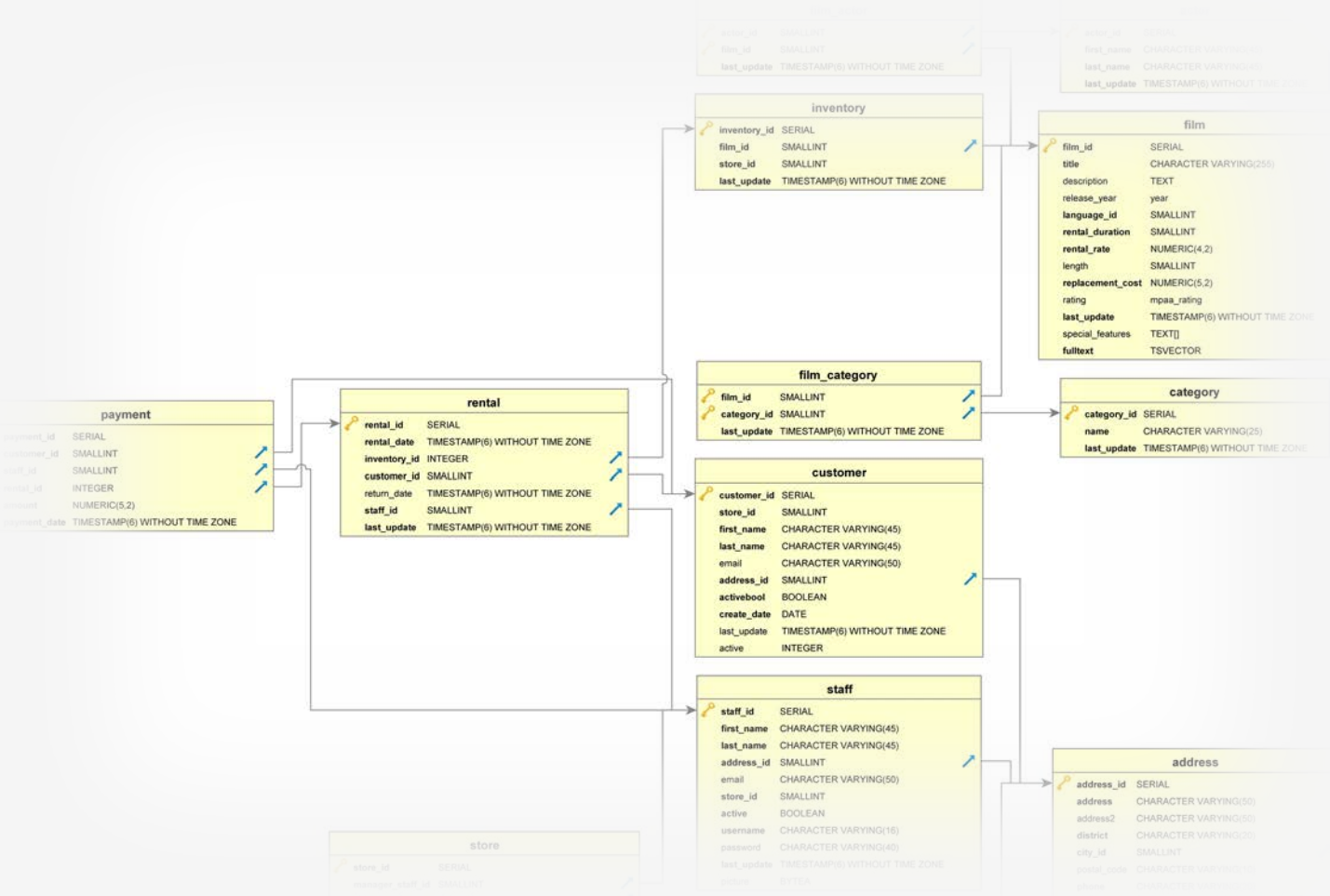
Technical information

- Language: SQL
- Software used: PostgreSQL 12
- Data base contained information about Rockbuster's film inventory, customers, payments, etc. Its size was around 3MB.

Answering business questions for an online video rental company

Tasks performed

- Loaded the data set into the PostgreSQL database
- Extracted entity relationship diagram (ERD) using DbVisualizer
- Migrated the database using Extract, Transform, and Load (ETL)
- Wrote SQL statements with the following effects:
  - ✓ Checking for duplicates and missing values
  - ✓ Joining tables
  - ✓ Performing subqueries
  - ✓ Using Common Table Expressions (CTE)
- Used a Development, Staging, and Production (DSP) model



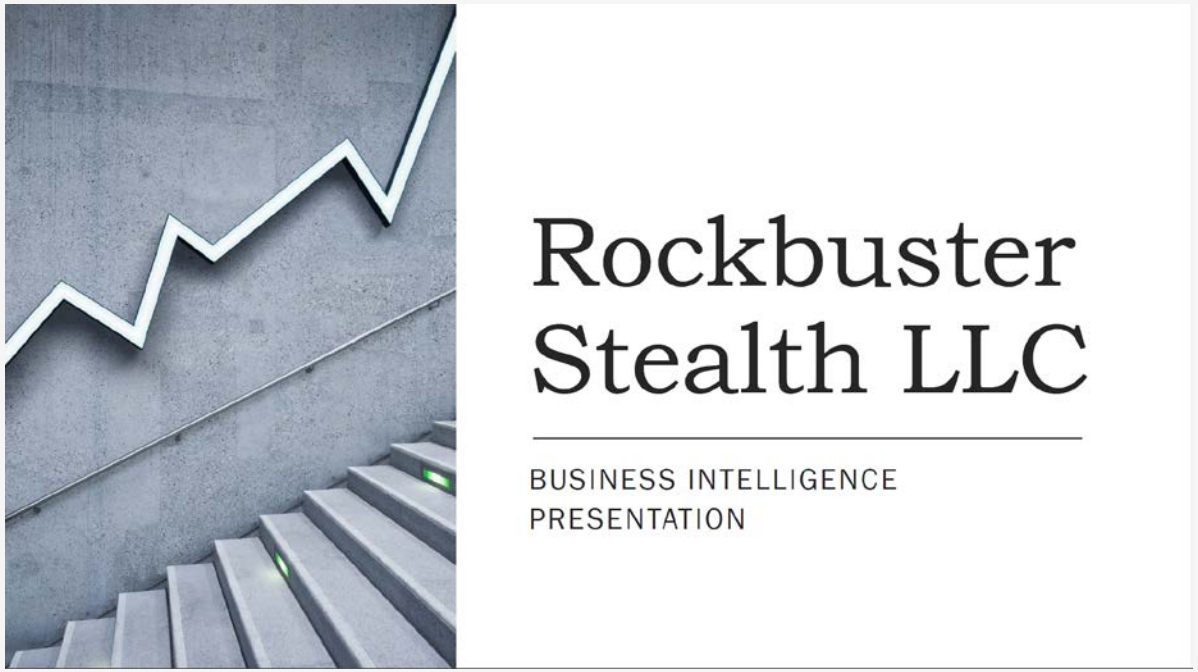
Answering business questions for an online video rental company

Results:

1. Created a data dictionary to serve as reference information defining the structure of the database

2. Created a final presentation to communicate the findings using Tableau Business Intelligence and Analytics Software

3. Created a technical document containing descriptive statistics, all the queries wrote during the project, as well as their results



Case-study 4

Marketing strategy for an online grocery store

Marketing strategy for an online grocery store

Project Overview

Instacart is an online grocery store that operates through an app. The project's objective was to increase sales by uncovering more information about the sales patterns and the clients' purchasing behaviours. My task was to perform an initial data and exploratory analysis of the data to derive insights and suggest strategies for better segmentation..

Key Questions

- What are the busiest days of the week and hours of the day?
- What are the times of the day when people spend the most money?
- Which departments have the highest frequency of product orders?
- What classifications does the demographic information suggest?
- What differences can you find in ordering habits of different customer profiles?

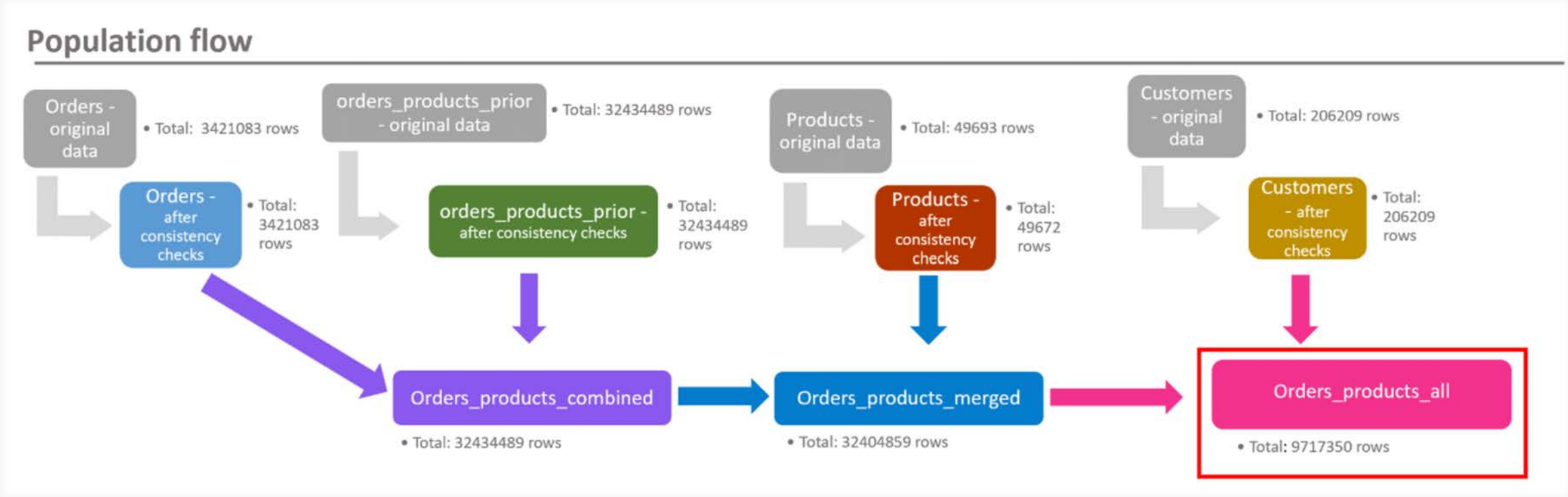
Technical information

- Language used: Python and relevant libraries (pandas, NumPy, os, matplotlib, scipy, and seaborn).
- Software used: Anaconda

Marketing strategy for an online grocery store

Information about the data base

For this project, I used four open-source data sets from Instacart. The following diagram indicates the characteristics of the original data sets and explain how the data sets were combined during the project.

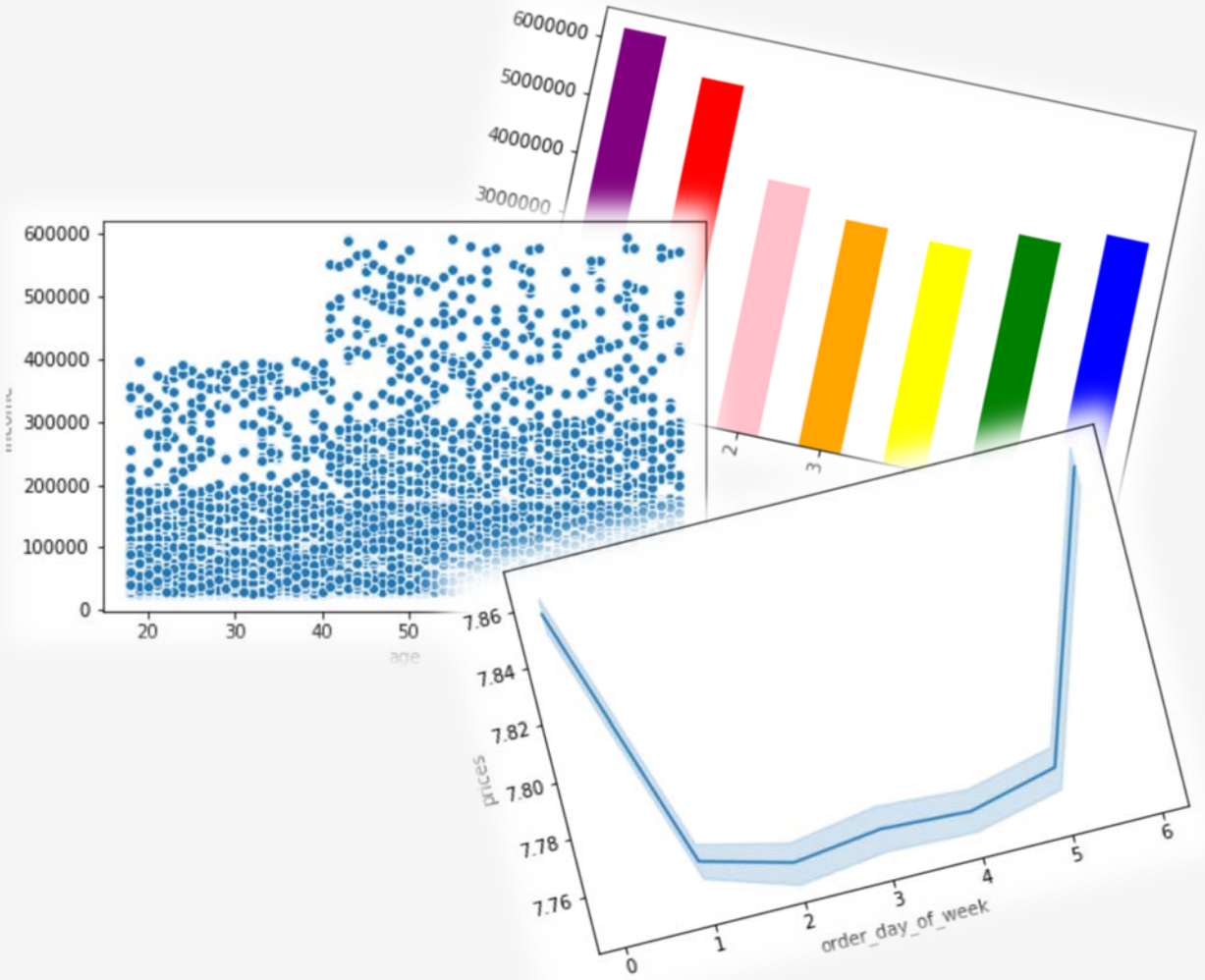


The grey boxes in the first row of the populatin flow represent the four original data sets I used for this assignment. The second row of boxes represents the data sets after manipulation, e.g. removal of missing values and duplicates. The third row represents the merges that were performed between the datasets.

Marketing strategy for an online grocery store

Tasks performed

- Successfully installed all required libraries.
- Conducted descriptive checks after importation of data.
- Organized project folders in logic and intuitive manner.
- Performed data consistency checks and addressed all cases of duplicate data, missing data, and mixed-type columns.
- Merged all project data into a single data set.
- Derived new variables, using if-statements, user-defined functions, the loc() function, and for-loops.
- Used the groupby() function to create flags and summary columns of descriptive statistics.
- Created data visualizations, such as histograms, bar charts, line charts, and scatterplots, to communicate insights to stakeholders.



Marketing strategy for an online grocery store

Results:

Created a final report containing:

- ✓ Newly created customer profiles.
- ✓ Detailed analysis of order behaviour based on different customer groups.
- ✓ A summary of analysis findings and a description of connections found in the data.
- ✓ A description of the methodology used in the analysis.
- ✓ Recommendations for Instacart stakeholders

Case-study 5

Anti-money laundering projects at a global bank

Anti-money laundering projects at a global bank

Project Overview

As a data analyst working for a well-known global bank, my job is to provide analytical support to its anti-money-laundering compliance department. This will involve a variety of data-related projects to help the bank assess client risk and transaction risk, as well as reporting on metrics.

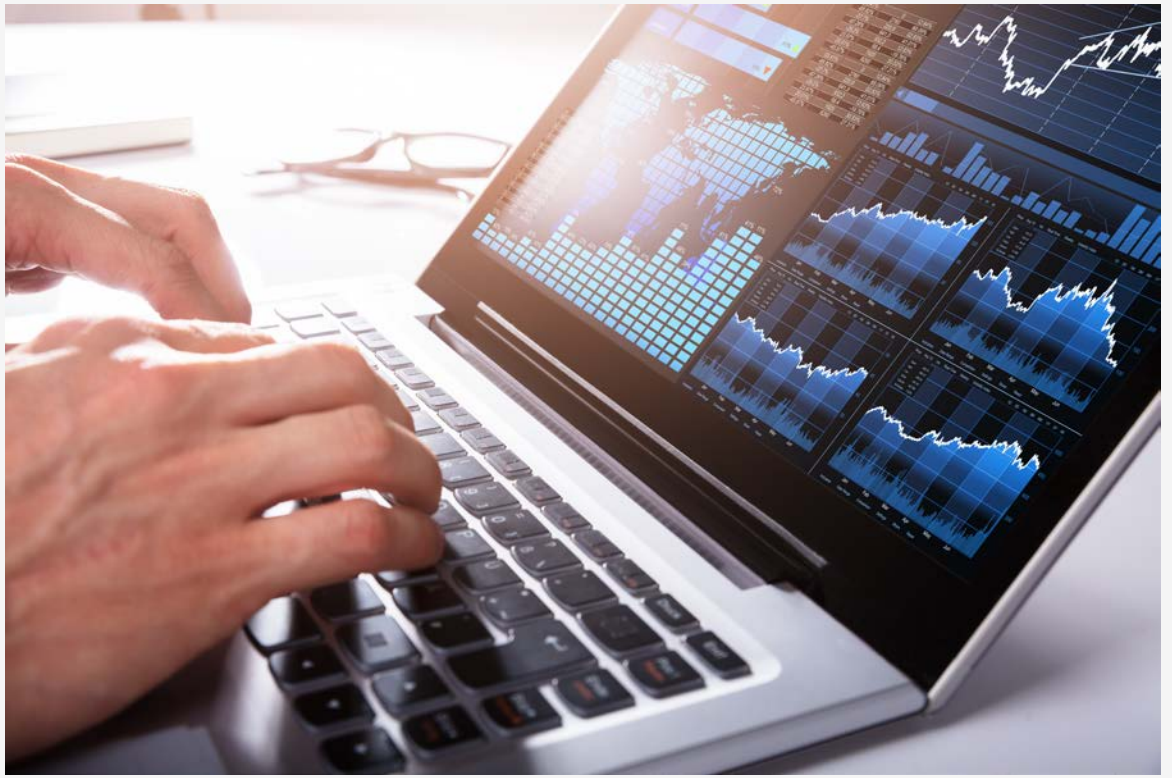
Key Objectives

- Learn how to handle big data to extract knowledge from it.
- Understand data bias and ethics issues around data.
- Learn to apply data mining.
- Discuss predictive analysis and models such as linear regression.
- Use time-series analysis and time-series forecasting.
- Use GitHub to learn, share and communicate data skills.

Anti-money laundering projects at a global bank

Tasks performed

- Learned about big data and Apache Hadoop.
- Practiced the identification of bias and ethical issues as a data analyst.
- Carried out steps in the data mining process.
- Created a decision tree model to test the outcomes of an analysis.
- Differentiate between regression and classification models.
- Analysed the output of a linear regression.
- Worked with time-series and moving averages.
- Research different time forecasting models.
- Created a GitHub repositories to host my SQL and Python work.



Answering business questions for an online video rental company

Results:

1. Created this storyline to showcase my data projects.

2. Created a video presentation of myself discussing data concepts to serve as job interview practice.